

of

RICHARD HAMMOND

and

MATTHEW CURRIE

for

GATE TECHNOLOGY FOR STRAINED SURFACE CHANNEL AND STRAINED BURIED CHANNEL MOSFET DEVICES

GATE TECHNOLOGY FOR STRAINED SURFACE CHANNEL AND STRAINED BURIED CHANNEL MOSFET DEVICES

PRIORITY INFORMATION

5 This application claims priority from provisional application Ser. No. 60/223,595 filed
August 7, 2000.

BACKGROUND OF THE INVENTION

10 The invention relates to gate technology for strained surface channel and strained buried
channel MOSFET devices.

15 The advent of high quality relaxed SiGe layers on Si has resulted in the demonstration
of field effect transistors (FETs) with carrier channels enhanced via strain. The strain can be
incorporated in the channel due to the lattice mismatch between the channel and the relaxed
SiGe created by a change in the Ge concentration between the channel layer and the relaxed
SiGe layer. For example, a Ge concentration of 20% Ge in the relaxed buffer is high enough
such that a thin strained Si layer can exhibit electron mobilities as high as 1000-2900 cm²/V-
sec. Also, if the Ge concentration in the channel is greater than the concentration in the
buffer, hole channel mobilities can be enhanced. For example, a relaxed buffer concentration
of 60-70% Ge can compressively strain a Ge channel layer, creating potentially extremely high
20 hole mobilities.

 Although the exact physics of carrier scattering are not known inside short-channel
FETs, one thing is clear: these enhanced mobilities translate into increased device
performance, even at very short gate lengths. In addition to higher speed and a different

power-delay product, the use of strained channels allows for the incorporation of new FET structures into Si-based circuits. Thus, it is anticipated that the high performance, new flexibility in device design, and economics of using a Si-based platform will lead to a plethora of new circuits and products.

5 With regards to these new circuits and products, the devices based on metal-insulator-semiconductor (MIS) or metal-oxide-semiconductor (MOS) gate technology are the most intriguing, since these devices can follow very closely the processes already used in Si VLSI manufacturing. Two main types of devices are of particular interest: the surface channel device and the buried channel device, examples of which are shown in FIGs. 1A and 1B.

10 FIG. 1A is a cross section of a block diagram of a strained Si surface channel device 100, in which a thin strained Si layer 102 is grown atop a relaxed SiGe virtual substrate. The SiGe virtual substrate can be relaxed SiGe 104 on a SiGe graded buffer 105 (as shown in Figure 1a), relaxed SiGe directly on a Si substrate 106, or relaxed SiGe on an insulator such as SiO₂. The device also includes a SiO₂ layer 108 and gate material 110.

15 FIG. 1B is a cross section of a block diagram of a strained Si buried channel device 112, in which a SiGe layer 116 and a second strained Si layer 120 (used for gate oxidation) cap the strained Si channel layer 114. The structure also includes a graded SiGe buffer layer 125 and a second relaxed SiGe layer 126. In both device configurations, a gate oxide 122 is grown or deposited and the gate material 124 is deposited to form the (MOS) structure. Although
20 only devices with strained Si channels are shown in FIGs. 1A and 1B, the invention is applicable to any heterostructure device fabricated on a relaxed SiGe platform. For example, the heterostructure strained channel could be Ge or SiGe of a different Ge content from that of

the underlying SiGe virtual substrate. However, the following description will focus on the applicability of the invention to the strained Si device variants illustrated in FIGs. 1A and 1B.

In order to form the MOS gate of the heterostructure device, the SiGe would ideally be oxidized directly in the buried channel device, and the strained Si would be oxidized directly in the surface channel device. Unfortunately, there are problems due to the nature of the Si/SiGe heterostructures in both cases that render the direct oxidation process unsatisfactory.

First consider the surface channel device. Since Si is being oxidized, the interface state density at the resulting SiO₂/Si interface is low, and an electrically high quality interface results. However, all oxidation and cleaning processes during the device and circuit fabrication consume the Si material. In conventional Si processing, there is generally little worry about Si consumption since so little material is consumed compared to any limiting vertical dimension early in the fabrication process. However, in the case of the strained surface channel FET described here, the top strained Si layer is typically less than 300Å thick, and thus too much Si consumption during cleaning and oxidation steps will eliminate the high mobility channel.

One obvious solution is to simply deposit extra Si at the surface, planning for the removal of the Si that occurs during processing. However, the channel strain, which gives the channel its higher carrier mobility, limits the Si layer thickness. At a great enough thickness, the Si layer will begin to relax, introducing misfit dislocations at the Si/SiGe interface. This process of dislocation introduction has two deleterious effects on device performance. First, the strain in the Si is partially or completely relieved, potentially decreasing the carrier mobility enhancements. Second, dislocations can scatter carriers, decreasing carrier mobility.

Dislocations can also affect device yield, reliability, and performance.

The buried channel case appears to be a better situation at first, since the Si layer thickness is buried. However, in this case, direct oxidation of SiGe creates a very high interface state density at the oxide/SiGe interface, leading to poor device performance. A known solution in the field is to create a thin Si layer at the surface of the buried channel structure. In this structure, the surface layer is carefully oxidized to nearly consume the entire top Si layer. However, a thin layer of un-oxidized Si is left so that the interface to the oxide is the superior SiO₂/Si interface rather than the problematic oxide/SiGe interface. Although this sacrificial surface Si layer solves the interface electronic property issue, the structure now has the same limits as the structure described above, i.e., the sacrificial Si layer will be slowly etched away during Si processing, possibly leading to exposure of the SiGe and degradation of the electrical properties of the interface as described.

SUMMARY OF THE INVENTION

In accordance with the invention there is provided a method of fabricating a semiconductor device including providing a semiconductor heterostructure, the heterostructure having a relaxed Si_{1-x}Ge_x layer on a substrate, a strained channel layer on the relaxed Si_{1-x}Ge_x layer, and a Si_{1-y}Ge_y layer; removing the Si_{1-y}Ge_y layer; and providing a dielectric layer. The dielectric layer includes a gate dielectric of a MISFET. In alternative embodiments, the heterostructure includes a SiGe spacer layer and a Si layer.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGs. 1A and 1B are cross sections of block diagrams of strained Si surface and buried channel devices, respectively;

FIGs. 2A and 2B are cross sections of block diagrams of starting heterostructures for surface channel and buried channel strained MOS, respectively, in accordance with the invention;

FIGs. 3A-3D are block diagrams showing the process sequence for a strained surface channel MOS device;

FIGs. 4A-4D are block diagrams showing the process sequence utilizing the gate structure for a buried channel device;

FIG. 5 is a graph of oxidation rates, under a wet oxidation ambient at 700°C, of SiGe alloys, with Ge contents of 0.28 and 0.36, compared to the oxidation rate of bulk silicon;

FIG. 6 is a graph showing the oxide thickness of both a $\text{Si}_{0.7}\text{Ge}_{0.3}$ alloy and a Si/ $\text{Si}_{0.7}\text{Ge}_{0.3}$ heterostructure;

FIG. 7 is a cross-sectional transmission electron micrograph (XTEM) of the Si/ $\text{Si}_{0.7}\text{Ge}_{0.3}$ heterostructure;

FIG. 8 is a XTEM image of the identical Si/ $\text{Si}_{0.7}\text{Ge}_{0.3}$ heterostructure after wet oxidation followed by oxide removal via a wet etch;

FIG. 9 is a structure for a buried channel MOSFET using relaxed SiGe and strained Si in accordance with the invention; and

FIG. 10 is a graph showing a plot of the middle SiGe layer thickness (h_2) and the resulting misfit dislocation spacing.

DETAILED DESCRIPTION OF THE INVENTION

To eliminate the issue of losing valuable surface Si, an innovative step that has not been previously considered can be employed. In fact, any interest in this area is dominated by discussions of how to change the Si device and circuit process to conserve Si consumption. Although these are certainly possibilities, such constraints severely limit process flexibility, alter the process further from the conventional Si process, and most likely will increase the cost of the fabrication process.

A solution for the buried channel and surface channel structures is to actually deposit another SiGe layer after the desired device structure (which, in the buried channel heterostructure, includes the sacrificial Si layer for oxidation). The structures are shown in FIGs. 2A and 2B.

FIG. 2A is a cross section of a block diagram of a starting heterostructure 200 for surface channel strained MOS in accordance with the invention. The structure 200 includes a Si substrate 202, a SiGe graded buffer 204, a relaxed SiGe layer 206, and a strained-Si channel layer 208. FIG. 2B is a cross section of a block diagram of a starting heterostructure 214 for buried channel strained Si MOS. The structure 214 includes a Si substrate 216, a SiGe graded buffer 218, relaxed SiGe layers 220 and 230, a first strained-Si channel layer 222 and a second strained-Si layer 224 for the gate oxide.

These structures are identical to those depicted in FIGs. 1A and 1B before the gate stack formation, except for the addition of a SiGe capping layer 210, 226 and an optional Si capping layer 212, 228. Since the SiGe layer 210, 226 is closely lattice-matched to the relaxed

SiGe layer below the device layers, there is essentially no limit on the thickness of the SiGe layer. This SiGe layer thickness can be tuned to the thickness of material removed before gate oxidation, so that the strained Si layer is exposed just before oxidation. Alternatively, the SiGe can be thicker than the removal thickness and then can be selectively removed. In fact, as described below, SiGe can be selectively removed with respect to Si using a variety of conventional Si-based processes. Therefore, cleaning and oxidation steps can be performed during the Si device and circuit fabrication process with little worry of consuming the precious strained Si and/or the sacrificial strained Si. One only needs to create a SiGe thick enough such that it is not totally consumed before the critical gate oxidation step.

An additional option can be to place yet another Si layer 212, 228 on top of the additional SiGe layer 210, 226. In some processing facilities, the idea of SiGe on the surface, instead of Si, is a factor for concern. In this case, another Si layer can be deposited on top of the additional SiGe layer described above. By choosing the Ge concentration in the additional SiGe layer to be greater than that of the virtual buffer, a compressive layer can be created; thus, if this additional optional Si layer is greater than the critical thickness, there is no possibility of dislocations moving into the device layers. This phenomenon occurs since the Si layers are tensile, and therefore dislocations introduced into the top optional Si layer have a Burgers vector that will not allow them to glide favorably in the compressive layer below. The dislocations in the top optional Si layer (if the Si layer critical thickness is exceeded) will not penetrate into the layers beneath it, and therefore as much Si can be deposited as desired. In fact, this optional Si capping layer need not be strained at all in this case and can serve as a protective sacrificial layer even if it is fully relaxed.

FIGs. 3A-3D are block diagrams showing the process sequence for a strained surface channel MOS device utilizing the gate structure described above (the process is shown for a structure without an optional strained surface layer). FIG. 3A shows the initial Si/SiGe heterostructure 200 shown in FIG. 2A. FIG. 3B shows the structure after the completion of the initial steps of a Si VLSI process, which could include wet chemical cleans and oxidation steps. Thus, in FIG. 3B, the protective SiGe capping layer 210 has been reduced in thickness, as a portion of the layer has been consumed during processing. Next, the remainder of the protective SiGe capping layer 210 is selectively removed, leaving the underlying Si layer 208 intact and exposed. A sacrificial oxidation step and oxide strip can also be performed at this point to improve the quality of the exposed Si surface.

The resulting structure is shown in FIG. 3C. FIG. 3D shows the final device structure after gate oxidation to form a gate oxide 300, a structure in which the minimum possible amount of Si was consumed prior to the gate oxidation step. Alternatively, at this point an alternate gate dielectric could be deposited on the exposed Si surface. A pristine Si surface is just as important for a high quality interface with many deposited gate dielectrics as it is for a thermally grown SiO₂ gate dielectric.

FIGs. 4A-4D are block diagrams showing the process sequence utilizing the gate structure for a buried channel device (the process is shown for a structure without an optional strained surface layer) using the initial Si/SiGe heterostructure 214 shown in FIG. 2B. The process steps are identical to those of FIGs. 3A-3D, but in the final heterostructure, the Si channel layer 222 is separated from the gate dielectric 400 by a SiGe spacer layer 220, thus forming a buried channel. Using selective processes to etch down to the buried Si channel or

the top Si layer can use the starting heterostructure 214 in FIG. 4A to form a surface channel device. Such a process can result in enhancement mode and depletion mode devices that can in turn be used to create E/D logic circuits as well as a plethora of analog circuits.

In both sequences, an exemplary sequence of steps is: 1. Pre-gate-oxidation cleaning
5 steps and oxidation; 2. Selective etch or oxidation to remove residual protective SiGe layer;
3. Sacrificial oxide formation on Si; 4. Sacrificial oxide strip; 5. Gate oxidation.

It will be appreciated that steps 3 and 4 can be optional, depending on whether there may be a small amount of Ge left on the surface after the selective removal of the SiGe protection layer. When the original heterostructure is grown, the SiGe/Si interface will not be
10 infinitely abrupt, and therefore it is possible to have a small amount of Ge in the optimally pure Si layer. A sacrificial oxide step can be employed to remove an additional small amount of the Si layer to ensure that pure Si is oxidized in the gate oxidation step, ensuring high quality gate oxide.

The second step, the selective removal of the residual SiGe protective material, can be
15 accomplished in a variety of ways. One convenient process is a wet oxidation step, preferably at 750°C or below. Under wet oxidation at these temperatures, SiGe is oxidized at rates that can be 100 times greater than rates oxidizing Si under the same conditions. Thus, in order to expose the Si for gate oxidation, one can simply do a wet oxidation of the SiGe layer and selectively stop at the Si layer. The oxidized SiGe can be stripped to expose the Si. It is important to note here that
20 the low temperature is not only important for the selectivity in the oxidation process, but also the low temperature is important to minimize or prevent the snow-plowing of Ge in front of the oxidation front, a known problem in the direct oxidation of SiGe.

FIG. 5 is a graph of oxidation rates, under a wet oxidation ambient at 700°C, of SiGe alloys, with Ge contents of 0.28 and 0.36, compared to the oxidation rate of bulk silicon. It is evident from the graph that, under such conditions, the oxidation rate of SiGe increases as the Ge content of the film increases.

FIG. 6 is a similar graph, showing the oxide thickness of both a $\text{Si}_{0.7}\text{Ge}_{0.3}$ alloy and a Si/ $\text{Si}_{0.7}\text{Ge}_{0.3}$ heterostructure. Again, the oxidation conditions were 700°C in a wet ambient; however, FIG. 6 depicts very short oxidation durations compared to FIG. 5. The Si/ $\text{Si}_{0.7}\text{Ge}_{0.3}$ heterostructure consists of a 50Å strained Si buried layer, followed by a 30Å $\text{Si}_{0.7}\text{Ge}_{0.3}$, a 20Å strained Si layer and finally a 50Å $\text{Si}_{0.7}\text{Ge}_{0.3}$ capping layer.

A cross-sectional transmission electron micrograph (XTEM) of the Si/ $\text{Si}_{0.7}\text{Ge}_{0.3}$ heterostructure is shown in FIG. 7. It should be noted from FIG. 6 that the presence of strained Si layers in the heterostructure results in a dramatic retardation in the oxidation rate when compared to the oxidation rate of the uniform $\text{Si}_{0.7}\text{Ge}_{0.3}$. This retardation of the oxidation rate forms the basis of the selective removal of SiGe alloys over strained Si epitaxial layers.

FIG. 8 is a XTEM image of the identical Si/ $\text{Si}_{0.7}\text{Ge}_{0.3}$ heterostructure after wet oxidation at 700°C for 2 minutes followed by oxide removal via a wet etch. It is apparent that the thin strained Si layer is unaffected by the selective oxidation and remains fully intact. Based on the data shown in FIG. 5, an oxidation duration of 2 minutes far exceeds that required to fully oxidize the 50Å $\text{Si}_{0.7}\text{Ge}_{0.3}$ capping layer of the heterostructure. The very thin dark band, which is apparent on the surface of the strained Si layer, is a snow-plowed high Ge content layer that occurs during oxidation. Such a layer may be removed using a simple chemical clean or a

sacrificial oxidation step, either or both of which typically occur prior to the formation of the gate oxide.

Alternatively, the protective SiGe capping layer can be removed via selective dry or wet chemical etching techniques. For example, at high pressures ($> 200\text{mT}$) and low powers, CF_4 dry etch chemistries will etch relaxed SiGe films with high selectivity to Si. Mixtures of hydrofluoric acid (HF), hydrogen peroxide (H_2O_2), and acetic acid (CH_3COOH) will also selectively etch relaxed SiGe layers over Si at selectivities of 300:1 or more. Other potential selective wet chemical mixtures include HF, water (H_2O), and either H_2O_2 or nitric acid (HNO_3).

Additionally, the stability of the entire structure can be improved by increasing the Ge concentration in the intermediate SiGe layer, and also the top SiGe layer if desired. Below, energetic calculations are used to reveal a guide to creating semiconductor layer structures that increase stability with respect to misfit dislocation introduction.

The critical thickness for a buried channel MOSFET using relaxed SiGe and strained Si has been determined using the energy-balance formulation. The structure considered is the one shown in FIG. 9. The structure 900 includes a 30% SiGe virtual substrate 902 topped by a 80\AA strained Si layer 904, a SiGe layer with Ge concentration x_2 and thickness h_2 906, and an additional 30\AA of strained Si 908. Additional stability would result from the addition of an additional SiGe cap layer as described previously. To simplify, the example of FIG. 9 considers only the increased stability created by increasing the Ge concentration (x_2) or thickness (h_2) of the SiGe intermediate layer. Additionally, since the SiGe cap layer is removed during processing, the stability of the heterostructure with the SiGe cap removed is or

primary importance.

In device processing, one must consider the critical thickness of the entire structure with respect to the relaxed virtual substrate. Individual layers that exceed the individual critical thicknesses are not explicitly ruled out, so one practicing the art would have to verify that none of the layers that are introduced into the desired structure exceed the individual layer critical thicknesses. In other words, in the following calculation it is assumed that each layer in the structure is below its critical thickness with respect to the relaxed buffer.

One key to the formulation is to realize that this calculation should be done with respect to the plastic deformation of the layer composite, δ . Then, the dislocation array energy is the same expression regardless of the layer structure. The elastic energy in the individual layers is changed because of δ . In tensile layers, the strain is lowered by δ . In compressive layers, the energy is raised by δ .

Thus, the energy for a dislocation array (per unit area) inserted at the base of the composite is:

$$E_{\delta} = 2\delta D(1 - \nu \cos \alpha) [\ln(h_T/b) + 1]$$

where h_T is the total thickness of the composite ($h_1 + h_2 + h_3$), α is the angle between the dislocation line and the Burgers vector b , ν is the Poisson ratio, and D is the average shear modulus for a dislocation lying at the interface between the virtual substrate and the composite structure.

The total elastic energy (per unit area) in all the layers is:

$$E_e = \sum_{i=1}^3 Y \epsilon_i^2 h_i$$

where Y is the Young's modulus. Thus, the total energy of the system is:

$$E_T = E_\delta + E_\epsilon.$$

The energy can now be minimized with respect to δ (if the energy is lowest with no dislocations, then δ will have a less than or equal to zero value). The value of plastic

5 deformation then is (for the 3 layer example):

$$\delta = \frac{f_1 h_1}{h_T} - \frac{f_2 h_2}{h_T} + \frac{f_3 h_3}{h_T} - \frac{D(1 - \nu \cos^2 \alpha) \left[\ln \left(\frac{h_T}{b} \right) + 1 \right]}{Y h_T}$$

The examination of this solution reveals that a general formulation for any structure would be (for any structure of n layers):

$$\delta = \sum_i^n \frac{f_i h_i}{h_T} - \frac{D(1 - \nu \cos^2 \alpha) \left[\ln \left(\frac{h_T}{b} \right) + 1 \right]}{Y h_T}$$

10 where f has been assigned a negative value for compressive layers and positive value for tensile layers, and h_T is the total thickness of the structure:

$$h_T = \sum_i^n h_i.$$

The amount of plastic deformation and resulting misfit dislocation spacing S was calculated for the structure depicted in FIG. 9 as follows:

15 Lower strained Si layer thickness $h_1 = 80 \text{ \AA}$

Upper strained Si layer thickness $h_3 = 30 \text{ \AA}$

Middle SiGe layer thickness h_2 variable

Middle SiGe layer Ge concentration x_2 variable

Virtual substrate Ge concentration: 30%

FIG. 10 is a graph showing a plot of the middle SiGe layer thickness (h_2) and the resulting misfit dislocation spacing. The sharp upturn on the plots represents the critical thickness h_2 of the middle SiGe layer when the entire composite structure destabilizes and introduces dislocations at the channel/virtual buffer interface. The different curves are for the different compositions in the second layer h_2 . Very small increases in Ge result in a large jump in stability of the device layers. This suggests that it is possible to stabilize the layer significantly but not have the band structure altered that much. Adding an extra 5-10% Ge into the h_2 layer increases the stability drastically. For example, FIG. 10 indicates that over 100Å of 30% Ge is required to provide the stability of a 20Å layer of 45% Ge content.

Increasing h_2 even when the h_2 layer is lattice-matched to the virtual buffer increases the stability of the multilayer structure. In the equations above, the effect can be seen to be much weaker than when a compressive strain in h_2 is created. When f_2 is zero due to lattice matching to the virtual buffer, the increased stability with increasing h_2 comes from the fact that h_1 is increasing and therefore decreasing δ (and increasing S).

It will be appreciated that all the calculations are equilibrium calculations, and as usual, one might suspect that these numbers are somewhat conservative, although also consider that the layers possess many threading dislocations that can bend over at the critical thickness, so there are plenty of sources for misfit dislocation generation.

Sacrificial SiGe capping layers provide an innovative method for the protection of thin strained device layers during processing. Such layers shield these critically important strained channel layers from process steps, such as wet chemical cleans and oxidations, which consume

surface material. Before the growth or deposition of the gate dielectric, these protective SiGe layers can be selectively removed by standard processes such as oxidation or wet etching, revealing the intact strained device layer. Also presented is a guideline for engineering strained layer stacks such that relaxation via misfit dislocation is prevented. Compressively strained intermediate layers increase the stability of tensile channel layers, and also serve as a barrier for misfit dislocation introduction into the underlying layers.

Although the present invention has been shown and described with respect to several preferred embodiments thereof, various changes, omissions and additions to the form and detail thereof, may be made therein, without departing from the spirit and scope of the invention.

What is claimed is: